Widespread genetic exchange among terrestrial bacteriophages

Olin K. Silander*†‡§, Daniel M. Weinreich†¶, Kevin M. Wright*‖, Kara J. O'Keefe**, Camilla U. Rang*, Paul E. Turner**, and Lin Chao*

*Department of Biological Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093; ¶Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138; and **Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06520

Edited by Peter Palese, Mount Sinai School of Medicine, New York, NY, and approved October 27, 2005 (received for review April 15, 2005)

Bacteriophages are the most numerous entities in the biosphere. Despite this numerical dominance, the genetic structure of bacteriophage populations is poorly understood. Here, we present a biogeography study involving 25 previously undescribed bacteriophages from the Cystoviridae clade, a group characterized by a dsRNA genome divided into three segments. Previous laboratory manipulation has shown that, when multiple Cystoviruses infect a single host cell, they undergo (*i*) rare intrasegment recombination events and (*ii*) frequent genetic reassortment between segments. Analyzing linkage disequilibrium (LD) within segments, we find no significant evidence of intrasegment recombination in wild populations, consistent with (*i*). An extensive analysis of LD between segments supports frequent reassortment, on a time scale similar to the genomic mutation rate. The absence of LD within this group of phages is consistent with expectations for a completely sexual population, despite the fact that some segments have >50% nucleotide divergence at 4-fold degenerate sites. This extraordinary rate of genetic exchange between highly unrelated individuals is unprecedented in any taxa. We discuss our results in light of the biological species concept applied to viruses.

sex | virus | linkage disequilibrium | Cystoviridae | biogeography

Microbes are the most numerous entities in the biosphere, and viruses that infect bacteria (bacteriophages) constitute the majority of these organisms (1). Various mechanisms exist for genetic exchange (sex) between individual microbes, but the importance of these processes in structuring virus populations remains unclear. Several studies have estimated that the rate of genetic exchange in virus populations is relatively low (2–16). However, only a few medically or agriculturally important viruses have been examined, few of which represent random population samples. These biases prevent the drawing of general conclusions. Bacteriophages are an excellent choice for studying the rate and significance of genetic exchange in virus populations. It is increasingly evident that phages often exert significant control over the population sizes of their hosts and, thus, influence large-scale ecological and biogeochemical processes attributed to bacteria (17, 18). Furthermore, it has long been proposed that sex may be generally favored, because it promotes linkage equilibrium, which (in addition to *de novo* mutations) can provide the genetic variability that is the raw material for natural selection (19–23). Therefore, from a theoretical standpoint, it is crucial to gauge the rate of genetic exchange in phages and to determine whether sexual processes act to structure biological populations that comprise a major portion of our global ecosystem.

Bacteriophages from the Cystoviridae clade were first isolated in 1971 from bean straw infested with *Pseudomonas syringae* pathovar *phaseolicola* (*Pp*) (24). Cystoviridae are characterized by a tripartite dsRNA genome, and the three RNA segments per phage particle are referred to as L, M, and S (large, medium, and small, respectively). Additional phages from this clade were not isolated until 25 years later from the leaves of several agricultural species (25). Experimental manipulations have shown that intrasegment recombination in Cystoviruses is infrequent, occurring at a rate of $\approx 10^{-7}$ per segment per generation (26), but reassortment (exchange of segments between viruses) readily occurs when host cells are multiply infected (27). These characteristics are biologically similar to those of segmented RNA viruses of medical and agricultural importance, such as Influenza-A. Reassortment between very distantly related phages in the Cystoviridae clade has been documented *in vitro* (28). However, the frequency of reassortment among Cystoviruses in the wild is unknown.

Here, we report the results of a biogeography study involving Cystoviridae isolates from different regions across the United States. We use nucleotide sequence data derived from newly and previously isolated strains (25, 29) to characterize migration rates and linkage disequilibrium (LD); LD is a measure of the nonrandomness of the association between alleles at different sites. We test for a correlation between genetic and geographic distance between segments and for LD within and between segments. The rarity of intrasegment recombination documented in laboratory studies of the phage suggests that LD within segments will be high. In contrast, frequent genetic reassortment among segments would cause low LD between segments. We test for LD using three methods: a Mantel test of genetic association, a test using standard metrics of LD (r^2 and D'), and a third method based on measures of phylogenetic congruence. In contrast to previous studies of viral populations, we find no consistent evidence for LD between most segment pairs, despite the fact that some segments have >50% nucleotide divergence at 4-fold degenerate sites. This extraordinary rate of genetic exchange between highly unrelated individuals is unprecedented in any taxa and implies that sex within this group of viruses occurs frequently and, in some cases, without limitation by selection.

Materials and Methods

Sampling. Twenty-five phages were isolated from white clovers (*Trifolium repens*) on the campus of the University of California, San Diego (UCSD) and Chico, CA, and one from green bean (*Phaseolus vulgaris*) in Hamden, CT. General nomenclature is as

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: LD, linkage disequilibrium; *Pp, Pseudomonas syringae* pathovar *phaseolicola*.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ273591–DQ273662).

†O.K.S. and D.M.W. contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: olin.silander@env.ethz.ch.

§Present address: Eidgenössische Technische Hochschule Zurich, Ecology and Evolution, ETH-Zentrum NW 8092, Zurich, Switzerland.

‖Present address: Department of Biology, P.O. Box 90338, Duke University, Durham, NC 27708.

follows: state of origin, initials of collector, and order of sample isolation. Clones taken from a single clover are identified with an alphabetical suffix that is serially assigned. At the UCSD location, five clones were isolated from each of three clovers (CA/KW064a-e, CA/KW065a-e, and CA/KW066a-e) and five other phages (CA/KW067–071) were isolated from individual clovers, all lying within a 1 m^2 quadrant. Four phages (CA/KW051, 052, 072, and 073) were isolated from clovers located ≈30 m, 50 m, 375 m, and 600 m from this quadrant. CA/OS001 was isolated in Chico, CA, (865 km from San Diego), and CT/KO001 was isolated in Hamden, CT (3,995 km from San Diego). All of the CA/KW clones were collected in August and September of 2002, the CA/OS clone was collected in July of 2003, and the CT/KO clone was collected in June of 2003.

Culture. Individual clovers collected in CA were placed in 10 ml of LB broth (Lennox medium, pH 7.5) inoculated with 10^8 *Pp* cells and grown overnight at 25°C. This culture was filter-sterilized through a 0.22-μm membrane, and a volume was plated on a lawn of *Pp*. One to five plaques were isolated from each plate. During culture, maximum phage densities after overnight culture were measured at 10^5 per ml, and bacterial densities were maintained at $>10^7$ per ml; thus, the likelihood of reassortment during isolation was low (probability of coinfection for each individual isolate $<10^{-2}$). Phage isolation from plant material collected in CT differed, in that the initial medium was not inoculated with *Pp*.

Extraction, PCR, and Sequencing. Genomic extraction was performed as described in ref. 30 or by using QIAamp Viral RNA minikits (Qiagen, Valencia, CA). The dsRNA genome was reverse-transcribed by using random hexamers, and resultant cDNA was used as PCR template (primers available upon request). PCR products were purified by using the QIAquick PCR purification kit (Qiagen) or ExoSAP-IT (United States Biological, Swampscot, MA). Sequencing was performed by using the BigDye Terminator reaction v3.1 on an ABI 3100 (Applied Biosystems). Sequence reads were curated by eye by using the program SEQUENCHER. Isolate CA/KW066e could not be amplified and was not included in the analysis. All sequences used in this study have been deposited in GenBank (L segment, DQ273591–DQ273614; M segment, DQ273615–DQ273638; and S segment, DQ273639–DQ273662).

Population Structure Analysis. Genetic isolation by geographic distance for each segment was tested by a Mantel test (31) by using the program ZT (32). Genetic distances were calculated by using the general time-reversible model (33) with γ distributed rate variation among sites, as implemented in the program TREE-PUZZLE (34). Geographic coordinates were measured on the UCSD campus by GPS or downloaded from the U.S. Geological Survey website. Distances were measured by using a campus map or calculated by using the javascript program available at www.wcrl.ars.usda.gov/cec/java/lat-long.htm. Permutation tests to assess significance of LD within segments were performed by using the program PAIRWISE, part of the LDHat package (downloaded from www.stats.ox.ac.uk/~mcvean/LDhat/LDhat1.0/LDhat1.0.html). All other analyses of LD and other permutation tests were implemented in ACTIVEPERL 5.8.4.810.

The two metrics of LD we used were r^2 and D'. r^2 is the squared correlation coefficient (35) between loci and is a simple measure indicating how often an allele found at one locus is associated with an allele at a second locus. Consider a locus with two alleles, A and a. If there is a second locus with alleles B and b, and the frequency of each allele is denoted as p_i and the frequency of each genotype as p_{ij}, then $r^2 = (p_{AB} - p_A p_B)^2 / p_A p_B p_a p_b$.

D' is a measure of LD relative to the maximum possible value under the observed allele frequencies (36). Considering the above set of biallelic loci, $D' = D/D_{max}$, where $D \equiv p_{AB} - p_A p_B$ and D_{max} is the lesser of $p_A p_b$ or $p_a p_B$, if D is negative, or the lesser of $p_A p_B$ or $p_a p_b$, if D is positive. The sign of D' is arbitrary, so we use $|D'|$ in all cases. r^2 is sensitive to variation in allele frequencies between loci, whereas D' is less so (37).

For the analysis of LD between segments, r^2 or D' was measured between all pairs of polymorphic loci (nucleotides) and an average value was obtained for all nucelotides between each pair of segments. One set of segments was then randomized, and the average r^2 or D' was recalculated. This process was repeated 10,000 times to test whether the observed average LD statistic was significantly greater than expected for a random (unlinked) collection of segments.

Phylogenetic Analysis. Orthologous sequences were obtained from the National Center for Biotechnology Information for the L, M, and S segments of the Cystovirus phage ϕ6 (38–40), and the M and S segments from the Cystovirus phages ϕ7, ϕ9, and ϕ10 (25). No orthologous L-segment sequence is available for these latter three phages. Nucleotide sequences were aligned by using the program CLUSTALX (41) and manually corrected by eye. The program MRBAYES v3.0B4 (42) was used for phylogeny construction. Analyses were performed by using the general time-reversible model, with γ distributed rate variation across sites. MCMC chains were run for 200,000 generations, with the first 100,000 discarded as burn-in for construction of the 50% majority-rule consensus trees. Tree-to-tree distance metrics were calculated by using the program COMPONENT v2.00a (43). Five-hundred random trees generated by using COMPONENT or 500 random trees from the set of 5,000 most-recently visited trees of the MCMC chain were used for the phylogenetic comparisons. Two distance metrics were used: the number of identical and the number of different triplets between trees.

Results

Frequency of Cystoviridae Among All Phage Isolates. Of the plant material collected in CA, 40% of all phages that formed plaques on *Pp* had tripartite dsRNA genomes, identifying them as members of the Cystoviridae.

Phylogenetic Reconstruction. In total, the sequence data encompassed 413 bp of the L segment, 296 bp of the M segment, and 315 bp of the S segment for each isolate. These sequences were orthologous to base pairs 2,293–2,705, 70–362, and 240–554 in the L, M, and S segments, respectively. In the L segment, this range corresponds to the middle of the coding region of gene 2 (part of the RNA polymerase), whereas, in the M and S segments, it corresponds to part of the pac sequence and part of the first ORF. The pac sequence of each segment is important for segment packaging into the viral procapsid. We constructed Bayesian phylogenies using the sequenced portions of each genomic segment and the published sequences of four previously isolated Cystoviruses (25, 39, 40). The 50% majority-rule consensus trees are illustrated in Fig. 1. These phylogenies contain extensive polytomies, because many of the phage sequences are nearly identical. It is possible that the segments differing by 1–2 mutations were clonal before isolation. Approximately 20 phage generations occurred between isolation and sequencing, and, given a mutation rate of 10^{-5} per generation per nucleotide (44, 45), the probability of a single mutation among the 1,024 base pairs sequenced is ≈20%.

Segment divergence peaked at 19.2% (S segment), 33.2% (M segment), and 31.4% (L segment). At 4-fold degenerate sites within the L segment, divergence was >50%. The major portions of the S and M sequences were noncoding, so an analysis of polymorphism at 4-fold degenerate sites was not possible. In-

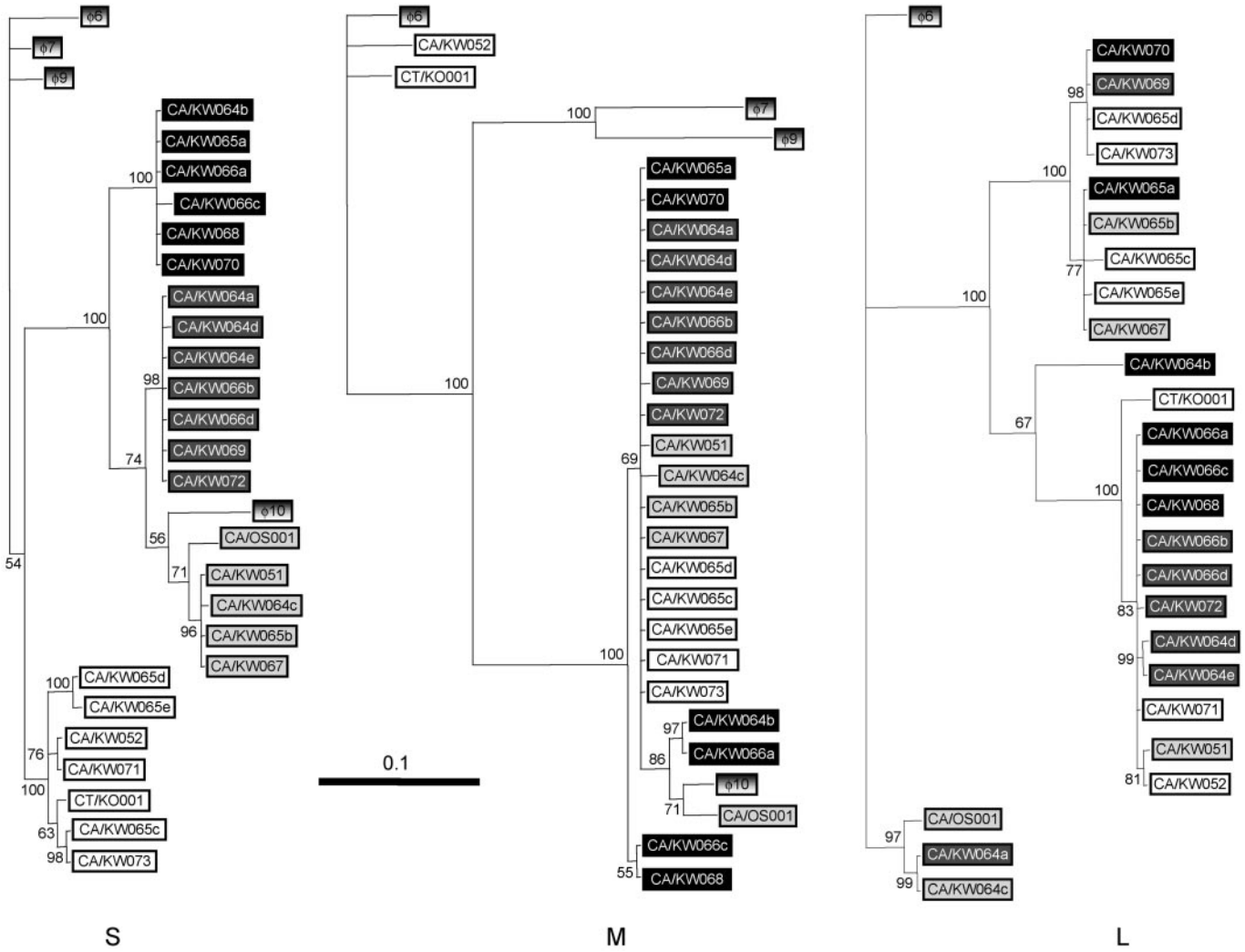

Fig. 1. The 50%-majority-rule consensus trees of the S, M, and L segments. Trees were constructed in MRBAYES by using the general time-reversible model, with γ distributed rate variation across sites. Posterior probabilities are indicated above each branch. The nomenclature of the isolates is as follows: state of origin, collector, and order of sample isolation; clones taken from a single clover are identified with an alphabetical suffix that is serially assigned. Gray shading indicates phylogenetic grouping based on clades within the small segment; previously isolated Cystoviridae are two-tone. The scale bar indicates 0.1 substitutions per site.

terestingly, the recently described Cystoviruses that grow on *Pp* (25) (isolated in 1999 in NY) fell within the M-segment diversity we observed, and two were slightly (with low posterior probability) more diverged on the S segment (Fig. 1). An orthologous sequence from the L segment for these clones is not available, preventing any phylogenetic comparisons. All of the environmental isolates were found to be at least an order-of-magnitude more diverged from ϕ6 than any other phage samples currently in our laboratory, ruling out laboratory contamination. Additionally, sequencing was done at two different laboratories: one in Cambridge, MA, and the other in San Diego. Finally, the phage isolation and sequencing done in San Diego were performed at two time periods approximately 1 year apart.

Population Structure. From the sequence data and phylogenetic reconstructions, it was apparent that migration and reassortment occurred frequently. Considering the three sets of multiple clones isolated from single clovers (14 viruses total), 2 viruses were identical for all three segments (CA/KW66b and CA/KW66d), and two viruses differed by two mutations (CA/KW64d and CA/KW64e). Each of the other 10 appeared to have undergone at least one reassortment event. In several other cases, phages that shared identical S or M segments had only distantly related L segments.

We tested all our isolates for an association between genetic and geographic distance by using a Mantel test. We removed from this analysis one of the two clonal isolates (CA/KW66b and CA/KW66d), taken from a single clover, because these phages were isolated from a single plate, and we therefore suspected that the two isolates may have been derived from a single progenitor phage. Including both of these isolates in the analysis did not qualitatively change the results. All other identical or nearly identical phages were isolated from different plates (e.g., CA/KW66c and CA/KW68), so they could not have been derived from a single progenitor phage. We also excluded previously described Cystoviruses (25), because they had been isolated at least 4 years previously, their geographic origin is unknown, and, for all of these phages, no orthologous L-segment sequence is available. Values of r, Pearson's correlation coefficient, together with the associated P values are shown in Table 1. No statistical evidence of geographic population structure was observed for the L or S segments, although a significant correlation between geographic and genetic distance was detected for the M segment.

EVOLUTION

Table 1. Results of Mantel tests of correlations between genetic and geographic distances and genetic distances between segments

Correlation	*r**	*P*†
S-geog	0.153	0.056
M-geog	0.532	0.014
L-geog	0.093	0.193
S–M	0.210	0.031
M–L	−0.030	0.528
S–L	−0.007	0.522

geog, geographic.
*Pearson's correlation coefficient.
†Proportion of permuted data sets with larger *r* values.

Intrasegment Recombination. We tested for intrasegment homologous recombination by examining the relationship between nucleotide distance and LD, using D′ and r^2 as metrics; recombination should manifest as a significant negative relationship between distance and either metric. In testing for recombination, we used all available sequences. We used the permutation test implemented in the PAIRWISE program from the LDHat package to test for significance of the relationship between D′ or r^2 and nucleotide distance. Previous experimental evidence has indicated that homologous recombination is extremely rare in Cystoviruses, occurring at a rate of $\approx 10^{-7}$ per segment per generation (26). The results here suggest that homologous recombination is also rare in natural populations. In two cases (L segment vs. D′, and M segment vs. r^2), there was a marginally significant negative relationship that did not remain after correcting for multiple tests (Table 2). Additionally, in both cases, the other metric held no significant relationship with nucleotide distance.

Reassortment. We also looked at levels of LD between segments. LD may exist between phage segments for three reasons: (*i*) Phages may not frequently migrate over large distances; (*ii*) if they do, coinfection and reassortment may be rare; and (*iii*) reassortants may have low fitness (on average) and thus experience negative selection. For all tests of LD between segments, we treated each segment as a single locus, because the above analysis suggested that intrasegment recombination is extremely rare or nonexistent. In the case of the Mantel test, LD was tested for by using the genetic distance between segments; for the r^2 and D′ metrics, it was done by averaging each metric over all pairs of nucleotides between segments to obtain a single measure of average LD between segments; for the phylogenetic test, we used phylogenetic congruity between segmental phylogenies. We performed the analyses using all of the recently isolated samples, except CA/KW066d, for the reasons discussed above.

Using a Mantel test, we looked for a positive correlation of genetic distance between segments, which would indicate LD.

Table 2. Results of permutation tests of correlations between LD metrics and nucleotide distance

Correlation	*r**	*P*†
S–D′	0.011	0.569
M–D′	0.089	0.992
L–D′	−0.069	0.045
S–r^2	0.005	0.519
M–r^2	−0.077	0.025
L–r^2	−0.004	0.430

*Pearson's correlation coefficient.
†Proportion of permuted data sets with smaller r values.

No association was found between the L segment and any other segment. A marginally significant correlation between S- and M-segment similarity was found ($P = 0.031$, Table 1).

We used a potentially more powerful test of LD in which we calculated the average value of r^2 and D′ over all polymorphic sites between segments. We used a permutation test to assess whether this value was significantly greater than the expected average r^2 or D′ if all segments were in linkage equilibrium. The majority of these tests were not significant. *P* values for the observed D′ values were 0.095, 0.318, and 0.261 for the S–M, M–L, and S–L segment comparisons, respectively (see Fig. 3, which is published as supporting information on the PNAS web site). *P* values for the observed r^2 values were 0.015, 0.337, and 0.306 for the S–M, M–L, and S–L comparisons, respectively (see Fig. 4, which is published as supporting information on the PNAS web site). In one instance (between the S and M segments), the observed LD metric was significantly higher than the observed metric between unlinked segments. However, this observation was not consistent for both measures of LD. The power of both this analysis and the Mantel test of LD may be compromised by homoplasy, which may obscure the true historical relationship between alleles.

In an attempt to account for this situation, we also used a phylogenetic analysis. Such an analysis would suggest LD between segments if the segment phylogenies were significantly more similar to each other than to a random set of phylogenies. Five-hundred random trees from the set of 5,000 most-recently visited trees in the MCMC chain were used for the analysis. The frequency of each tree in this data set is proportional to the probability that it is the true tree. This set of 500 trees was compared to two sets of other trees: five hundred of the most-recently visited MCMC trees from another segment, or 500 randomly constructed trees (43). We used two measures of tree-to-tree distance, the number of resolved identical triplets and the number of resolved different triplets between the sets of phylogenies (43).

By using the median number of identical triplets as an estimate of the true distance between phylogenies, we found that none of the segment phylogenies were significantly more closely related to each other than they were to a random set of trees (Fig. 2*A*, mean number of S–M identical triplets: 834, $P = 0.940$; M–L, 763, $P = 0.878$; S–L, 641, $P = 0.552$). However, there were significantly fewer different triplets between the S and M segment phylogenies (Fig. 2*B*, mean number of S–M different triplets, 1,119; $P = 0.047$; M–L, 1,143; $P = 0.074$; S–L, 1,214; $P = 0.211$).

Discussion

A tripartite dsRNA genome characterizes phages in the Cystoviridae clade, a historically depauperate group, with $\phi 6$ as its sole representative for >25 years after initial isolation. Eight other phages within this clade were recently isolated from bacteria-infested legumes (25), indicating that these viruses may be relatively common in many terrestrial habitats. Previous attempts to isolate additional Cystoviruses had failed, and it is unclear what has precipitated the more recent success. The high frequency of Cystoviruses within our environmental isolates hints that these phages may be a dominant predator in some bacterial communities.

Reassortment and Recombination. The most surprising result of our study is the complete lack of evidence for LD between some pairs of phage segments. LD is conspicuously absent between the S and L segments and the M and L segments, where we had the greatest power to detect it. Two factors decrease power in detecting LD: skewed allele frequencies (although this is mitigated by using randomization tests) and reduced genetic diversity. Thus, the two sets of segments in which we have the most

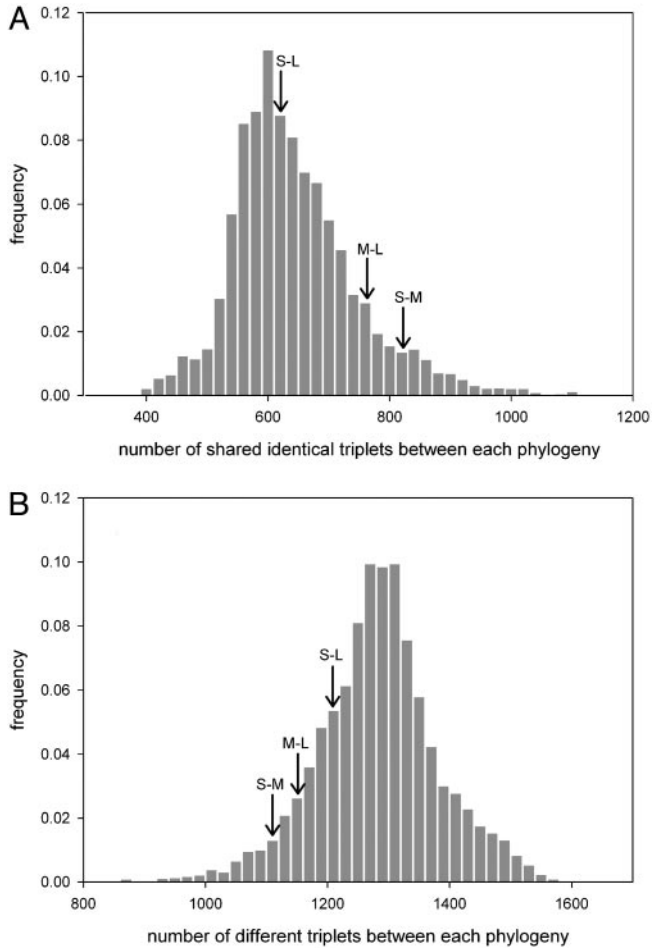

Fig. 2. Distributions of tree similarity metrics. (*A*) Number of resolved, shared identical triplets. Bars indicate the random expectation of identical triplets for 20,000 tree-to-tree comparisons. Arrows indicate the median number of identical triplets found between segment phylogenies, none of which showed statistical significance (see text). (*B*) Number of resolved different triplets. Bars indicate the random expectation of different triplets for 25,000 tree-to-tree comparisons. Arrows indicate the median number of different triplets between 500 segment phylogenies (25,000 comparisons); only the median S–M segment comparison was found to be statistically significant (see text).

power to detect LD are between the M and L, having the most genetic diversity; and between the S and L, having the least skewed allele frequencies. However, it is precisely these pairs that show no evidence of LD. These results thus imply that migration, coinfection, and selection all fail to limit reassortment between these segments. We consistently found LD (albeit at low levels) between the S and M segments; because all three LD tests bordered on significance, a lack of statistical power is likely to be important in the latter result. Overall, our study strongly suggests that the Cystoviridae experience frequent and widespread migration and regular coinfection of bacterial cells. Selection against reassortant phages seems to occur for distantly related S and M segments.

Several phages are identical for one segment and highly divergent for the other two (Fig. 1), indicating that the rate of reassortment is higher than the observed rate of substitution, even at 4-fold degenerate sites. If substitutions at 4-fold degenerate sites are neutral, we will observe them at the same rate as the per-base-pair mutation rate. We observed identical sequences for one segment and highly divergent sequences for a second segment; thus, reassortment has occurred before mutation at 4-fold degenerate sites within the first segment. Our sequence data encompassed a minimum of 7% of each segment, and, thus, we expect that the per-segment mutation rate does not exceed the per-segment rate of reassortment by >14 times (1/0.07).

Most virus studies concern strains pathogenic to humans and have concluded that reassortment occurs at low rates in virus populations. For example, reassortment in Influenza-A virus seems relatively uncommon (2–4) and depend on relatedness (6). Recent work has suggested that perhaps two or three reassortment events have occurred over a period of several years (4). Qualitative studies in Influenza-B virus have shown that, although reassortment is important for creating viral diversity (5, 9, 16), the rate does not approach that which we found in the Cystoviridae. High levels of LD are present between segments, even when examined across a relatively large time scale (25 years). Rotaviruses experience significant, although not high, rates of reassortment: Various measures of the frequency of reassortant viruses range between 2.7% (8) and 5.4% (14). Within the Bunyaviridae group, reassortment generally occurs infrequently (10) and depends on genetic similarity (7), whereas, in some taxa of pathogenic segmented viruses, reassortment has never been documented (15). Studies of disease-causing segmented viruses of plants have found low reassortment rates (13) and strong selection acting against reassortants (11, 12, 47). A drawback of these studies is that they give qualitative assessments of reassortment; however, the studies do suggest that it is not very common. It is thus surprising that our data indicate Cystovirus populations undergo genetic exchange at rates similar to those of obligately sexual populations. The result is even more surprising when one considers the amount of nucleotide divergence we observed between segments (>50% at 4-fold degenerate sites).

Laboratory assessments of intrasegment homologous recombination rates in the Cystoviridae (26, 46) are supported by our study. We found only marginally significant evidence for recombination; additionally, this evidence was inconsistent (Table 1). This test of intrasegment recombination (the decline of LD with increasing nucleotide distance) is extremely sensitive to small levels of recombination. Thus, although the result of the test is nonsignificant, its sensitivity suggests that recombination within segments is an extremely rare event. However, this analysis was done on a fairly small scale of hundreds of nucleotides; with sequence data spanning entire segments, significant levels of recombination may be detectable. It is important to note that the rate of recombination is extremely low relative to both the rate of mutation and the rate of reassortment ($\approx 10^6$-fold lower than both). Thus, homologous recombination is likely to play a relatively unimportant role in the evolution of this phage.

The primary limitation in the rate of homologous recombination is probably the nature of template replication in this phage family (26), although selection may also play a role in limiting recombination in various genomic regions. Specifically, some forms of epistasis will select against recombinants.

Genetic Exchange and the Viral Species Concept. Proper methods for assigning species boundaries in virus taxonomy are unclear, mostly because of inadequate understanding of what constitutes a species in naturally occurring viral populations. It would be informative to apply the biological species concept; viruses are capable of genetic exchange (2–16), and this method separates species according to barriers to genetic mixis. If viral species can be accurately delineated, effective population sizes can also be assessed. The population size of most natural viral populations is unknown, although this parameter may play a key role in processes such as disease emergence. It has been suggested that pathogens with higher mutation rates will produce more genetic variants and are, therefore, more likely to be generalists (48).

This same prediction applies to larger population sizes, because a more appropriate measure of genetic variation is the population mutation rate ($4N_e\mu$ in diploids). Thus, determination of viral species through the biological species concept and subsequent measurements of population size could provide an informative measure for the likelihood of emergent disease. It is likely that, in many viral groups, the rate of genetic mixis is a continuous trait, unlike the situation in most eukaryotic taxa, in which genetic exchange is, to a large extent, either present or absent, implying that, instead of distinct genetic clusters, such as those seen in eukaryotic and most prokaryotic taxa (49), loosely cohesive genetic clouds will be observed.

Migration. The geographic sampling was structured on a logarithmic scale, which should provide the most power for resolving structure at almost any scale. Notably, phages isolated from single clovers were not consistently more similar to each other than to phages isolated from sites across the country or from other previously isolated phages (Fig. 1). These data are thus consistent with frequent continent-wide migration in the Cystoviridae. To our knowledge, the only other study (50) addressing migration rates in phage populations also found continent-wide migration but in dsDNA (T7-like) marine phages. It is interesting that the only segment for which we find a significant association between geographical and genetic distance is the M segment. Limited migration cannot be the cause for this geographic structure, because neither the S nor L segments show evidence of regional structure. Nor is it likely that this structure has been introduced by a bias in the isolation process, because all of the samples should be subject to the same procedural bias in the laboratory. Rather, the result suggests that the M segment shows geographic structure because of selection. Notably, the M segment carries the genes responsible for host specificity (28); selection for local host adaptation may limit the amount of M segment migration.

Conclusions

Our data suggest that rates of genetic exchange can vary by orders of magnitude across viral taxa; although reassortment seems rather uncommon in many viruses, within the Cystoviridae, we find it to be commonplace. However, previous studies did not take a quantitative approach to examining rates of genetic exchange in wild populations of viruses. Therefore, our data should motivate greater effort to determine whether other viruses frequently undergo sex, which may cause them to feature a population structure that is truly panmictic. Mutation has long been appreciated for its role in promoting genetic variation in viruses (especially in those with RNA genomes) (44) and implicated as a key force in disease-related processes, such as the crossing of host–species boundaries (48, 51). But a dearth of studies that accurately measure rates of recombination or reassortment in natural populations might have caused the role of genetic exchange to be underappreciated. Our discovery that natural populations of highly diverged microbes can feature extremely low levels of LD is unprecedented.

We thank C. Burch (University of North Carolina, Chapel Hill) for PCR primers; J. Huelsenbeck and P. Andolfatto for insights into the phylogenetic and LD analyses, respectively; and P. Chisholm (Massachusetts Institute of Technology, Cambridge, MA) and D. Hartl (Harvard University) for laboratory space and reagents. This work was supported by a National Institutes of Health (NIH) Genetics Training Grant (to O.K.S.), NIH National Research Service Award GM207636 (to D.M.W.), a David Belkin Memorial Scholarship (to K.M.W.), National Science Foundation (NSF) Postdoctoral Fellowship DBI0400834 (to K.J.O.), and NSF Grant DEB0129089 (to P.E.T.).

1. Hendrix, R. W. (2002) *Theor. Popul. Biol.* **61,** 471–480.
2. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. (1992) *Microbiol. Rev.* **56,** 152–179.
3. Xu, X. Y., Lindstrom, S. E., Shaw, M. W., Smith, C. B., Hall, H. E., Mungall, B. A., Subbarao, K., Cox, N. J. & Klimov, A. (2004) *Virus Res.* **103,** 55–60.
4. Lindstrom, S. E., Cox, N. J. & Klimov, A. (2004) *Virology* **328,** 101–119.
5. Lin, Y. P., Gregory, V., Bennett, M. & Hay, A. (2004) *Virus Res.* **103,** 47–52.
6. Hatchette, T. F., Walker, D., Johnson, C., Baker, A., Pryor, S. P. & Webster, R. G. (2004) *J. Gen. Virol.* **85,** 2327–2337.
7. Henderson, W. W., Monroe, M. C., Jeor, S. C. S., Thayer, W. P., Rowe, J. E., Peters, C. J. & Nichol, S. T. (1995) *Virology* **214,** 602–610.
8. Iturriza-Gomara, M., Isherwood, B., Desselberger, U. & Gray, J. (2001) *J. Virol.* **75,** 3696–3705.
9. McCullers, J. A., Wang, G. C., He, S. Q. & Webster, R. G. (1999) *J. Virol.* **73,** 7343–7348.
10. Nemirov, K., Vapalahti, O., Lundkvist, A., Vasilenko, V., Golovljova, I., Plyusnina, A., Niemimaa, J., Laakkonen, J., Henttonen, H., Vaheri, A. & Plyusnin, A. (1999) *J. Gen. Virol.* **80,** 371–379.
11. Brown, J. K., Idris, A. M., Alteri, C. & Stenger, D. C. (2002) *Phytopathology* **92,** 734–742.
12. Garcia-Arenal, F., Fraile, A. & Malpica, J. M. (2001) *Annu. Rev. Phytopathol.* **39,** 157–186.
13. Miranda, G. J., Azzam, O. & Shirako, Y. (2000) *Virology* **266,** 26–32.
14. Watanabe, M., Nakagomi, T., Koshimura, Y. & Nakagomi, O. (2001) *Arch. Virol.* **146,** 557–570.
15. Vieth, S., Torda, A. E., Asper, M., Schmitz, H. & Gunther, S. (2004) *Virology* **318,** 153–168.
16. McCullers, J. A., Saito, T. & Iverson, A. R. (2004) *J. Virol.* **78,** 12817–12828.
17. Wommack, K. E., Hill, R. T., Kessel, M., Russekcohen, E. & Colwell, R. R. (1992) *Appl. Environ. Microbiol.* **58,** 2965–2970.
18. Fuhrman, J. A. (1999) *Nature* **399,** 541–548.
19. Kondrashov, A. S. (1993) *J. Hered.* **84,** 372–387.
20. Chao, L., Tran, T. R. & Matthews, C. (1992) *Evolution (Lawrence, Kans.)* **46,** 289–299.
21. Kondrashov, A. S. (1988) *Nature* **336,** 435–440.
22. Barton, N. H. & Charlesworth, B. (1998) *Science* **281,** 1986–1990.
23. Otto, S. P. & Lenormand, T. (2002) *Nat. Rev. Genet.* **3,** 252–261.
24. Vidaver, A. K., Koski, R. K. & Van Etten, J. L. (1973) *J. Virol.* **11,** 799–805.
25. Mindich, L., Qiao, X. Y., Qiao, J. A., Onodera, S., Romantschuk, M. & Hoogstraten, D. (1999) *J. Bacteriol.* **181,** 4505–4508.
26. Onodera, S., Sun, Y. & Mindich, L. (2001) *Virology* **286,** 113–118.
27. Mindich, L., Sinclair, J. F., Levine, D. & Cohen, J. (1976) *Virology* **75,** 218–223.
28. Qiao, X. Y., Qiao, J., Onodera, S. & Mindich, L. (2000) *Virology* **275,** 218–224.
29. Vidaver, A. K., Koski, R. K. & Van Etten, J. L. (1973) *J. Virol.* **11,** 799–805.
30. Day, L. A. & Mindich, L. (1980) *Virology* **103,** 376–385.
31. Mantel, N. A. (1963) *Cancer Res.* **27,** 209–220.
32. Bonnet, E. & Van de Peer, Y. (2002) *J. Stat. Software* **7,** 1–12.
33. Tavare, S. (1986) *Lect. Mathemat. Life Sci.* **17,** 57–84.
34. Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002) *Bioinformatics* **18,** 502–504.
35. Hill, W. G. & Robertson, A. (1968) *Theor. Popul. Biol.* **38,** 226–239.
36. Lewontin, R. C. (1964) *Genetics* **49,** 49–67.
37. Devlin, B. & Risch, N. (1995) *Genomics* **29,** 311–322.
38. Gottlieb, P., Metzger, S., Romantschuk, M., Carton, J., Strassman, J., Bamford, D. H., Kalkkinen, N. & Mindich, L. (1988) *Virology* **163,** 183–190.
39. Mcgraw, T., Mindich, L. & Frangione, B. (1986) *J. Virol.* **58,** 142–151.
40. Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D. H. & Kalkkinen, N. (1988) *J. Virol.* **62,** 1180–1185.
41. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
42. Huelsenbeck, J. P. & Ronquist, F. (2001) *Bioinformatics* **17,** 754–755.
43. Page, R. D. M. (1996) *Comput. Appl. Biosci.* **12,** 357–358.
44. Drake, J. W. & Holland, J. J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13910–13913.
45. Chao, L., Rang, C. U. & Wong, L. E. (2002) *J. Virol.* **76,** 3276–3281.
46. Mindich, L. (1996) *Semin. Virol.* **7,** 389–397.
47. Fraile, A., AlonsoPrados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & Garcia Arenal, F. (1997) *J. Virol.* **71,** 934–940.
48. Woolhouse, M. E. J., Taylor, L. H. & Haydon, D. T. (2001) *Science* **292,** 1109–1112.
49. Cohan, F. M. (2002) *Annu. Rev. Microbiol.* **56,** 457–487.
50. Breitbart, M., Miyake, J. H. & Rohwer, F. (2004) *FEMS Microbiol. Lett.* **236,** 249–256.
51. Moya, A., Holmes, E. C. & Gonzalez-Candelas, F. (2004) *Nat. Rev. Microbiol.* **2,** 279–288.